DicoPerec : pour une approche lexicographique de l'œuvre perecquienne

Montémont, Véronique

ATILF Nancy Université & CNRS
Institut universitaire de France

L'œuvre de Perec, depuis la mort de l'écrivain en 1982, connaît une fortune critique qui ne cesse de grandir : outre la parution d'une quinzaine d'ouvrages critiques consacrés à cet auteur, on dénombre plus d'une centaine d'articles, plusieurs colloques internationaux, un cycle de séminaires, une revue, *Le Cabinet d'amateur*, et une publication régulière, les *Cahiers Georges Perec*, qui compte à ce jour neuf numéros. Un fonds d'archives appartenant aux héritiers de l'écrivain et comportant ses manuscrits, a été déposé à la Bibliothèque de l'Arsenal par ces derniers, et la gestion en a été confiée à l'Association Georges Perec ; la mise à disposition de ce matériau documentaire a permis une diversification des approches critiques. Bernard Magné, pionner des études perecquiennes, a développé une recherche qui met, entre autres thèmes, l'accent sur le métatexte, le rôle générateur des contraintes et du nombre ; il a par ailleurs dégagé la notion fondatrice d'*aencrage*[i] (Magné, 1999). Claude Burgelin (Burgelin, 1990) , suivi par Matthieu Rémy (Rémy, 2003) privilégie une approche sociocritique, à laquelle invite le roman *Les Choses* (Perec, 1965), témoignage, entre autres dimensions, sur la société de consommation des années 60. La psychanalyse est une autre voie d'accès à l'œuvre (Burgelin, 1996), tout comme le biographique (Magoudi, 1996) ou encore l'étude spécifique de la judéité (Bénabou, 1984, Robin, 1993). La publication du *Cahier des charges de la vie mode d'emploi* (Hartje, Magné, Neefs, 1993) a ouvert la voie à une tradition d'études génétiques (Lejeune, 1991 ; Constantin, 2000, 2004) ; ont également été développées plusieurs études axées sur l'intermédialité : Perec et la photographie (Reggiani, 2001), Perec et le cinéma (De Bary et *al.*, 2006).

Si la thématique, pour des raisons liées à la formidable complexité des structures narratives et à la richesse du monde imaginaire de Perec, fascine, la relation à la langue, qui n'est certes pas oubliée par les perecquiens, est néanmoins abordée de manière plus dispersée : des études ont été consacrées à la relation de Perec au dictionnaire (Magné, 1997), aux procédures néologiques mises en œuvre dans les romans lipogrammatiques (Magné, 1989), à la phrase (De Bary, 2005), et même à la statistique lexicale (Hoyer, 1997). Mais il demeure difficile de proposer une vue d'ensemble de la stylistique et de la langue perecquienne : d'abord parce que son écriture, soumise à ce que l'Oulipo appelle des contraintes « dures » (le lipogramme en *e* de *La Disparition* apparaissant comme la plus spectaculaire), tend à varier pour cette raison de manière sensible d'un livre à l'autre. Ensuite parce que Perec a fait le choix d'une démarche enracinée dans le descriptif, éloignée, à première vue tout du moins, de tout excès d'ornementation littéraire, ce qui a pu, par exemple, inciter à le ranger au rang des écrivains pratiquant *l'écriture blanche*. Lui-même décrivait par ces mots sa démarche de remémoration dans *W* : « Je sais que ce je dis est blanc, est neutre, est signe une fois pour toutes d'un anéantissement une fois pour toutes » (Perec, 1975 : 59).

Cette neutralité revendiquée entre pourtant en tension, au sein de l'œuvre, avec une opulence lexicale tout à fait particulière, qui ne semble pas relever d'un goût pour le baroquisme ou d'un quelconque euphuisme. Comme le souligne (Magné, 1999 : 13), l'affrontement entre l'écrivain et la langue est autrement plus dur : Perec avait avoué sur le plateau d'une émission radiophonique que son ambition était de « [s]e mesurer avec les dictionnaires ». Dans ces conditions, on peut se demander dans quelle mesure la lexicographie ne serait pas susceptible de constituer une approche critique pertinente, voire novatrice, pour appréhender cette œuvre – impressionnante tant du point de vue quantitatif que du point de vue qualitatif – de manière holistique, au moyen de la rédaction d'un *Dictionnaire Perec*. Avant d'exposer ce

que sera la méthodologie de rédaction d'un tel outil, nous nous proposons d'étudier la problématique de que nous appellerons *le geste lexicographique* de Georges Perec car sa présence conditionne, dans une large mesure, le regard que nous pouvons porter sur la langue de cet auteur et la démarche scientifique qu'elle inspire.

1 Faits et gestes lexicographiques

Perec était un véritable passionné des dictionnaires : comme il le raconte dans *W ou le souvenir d'enfance* : « il y avait chez tante Berthe un grand dictionnaire Larousse en deux volumes. Peut-être est-ce là que j'ai appris à aimer les dictionnaires ». (Perec, 1975 : 201). L'inventaire de sa bibliothèque, effectué en 1983 par Catherine Binet et Eric Beaumatin, révèle qu'il en possédait plus d'une quarantaine : dictionnaires de langue (*Larousse du XX^e siècle* en 6 volumes, *Grand Larousse de la langue française* en 7 volumes, deux éditions du *Petit Larousse*, *Petit Robert*), dictionnaires de langue spécialisés (Pialat : *Larousse du Scrabble*, Nodier : *Dictionnaire des onomatopées*, Caradec : *Dictionnaire du français argotique et populaire*, Morier : *Dictionnaire de poétique et de rhétorique*, *Larousse des mots croisés*), dictionnaires techniques (Le Lionnais : *Dictionnaire des mathématiques*, Morvan : *Dictionnaire de* l'informatique, *Dictionnaire de la Seconde Guerre mondiale*), ainsi que de très nombreux recueils de mots croisés.

A ces collections lexicales préconstituées, il faut ajouter les listes que Perec a lui-même rédigées en vue de l'écriture de *La Disparition*, « micro-lexiques » (Magné, 1997 : 17) recensant les mots du français ne comportant pas la lettre *e*. L'auteur raconte : « vous avez donc pris un dictionnaire de langue française et vous avez recueilli tous les mots sans "e" » (Perec, 1972 : 4^e de couverture de la première édition). L'opération a ensuite donné lieu à des contre-listes, la monovalique en *e*, qui a servi à la rédaction des *Revenentes* (« vous avez donc récidivé en prenant, cette fois-ci, les mots ne comportant que la voyelle "e" ») ; ou encore la liste en *a* publiée dans *Ca m'intéresse* (Magné, 1997 : 15). Par ailleurs, plusieurs œuvres ayant pour thème la liste ou l'inventaire s'inscrivent dans la même dynamique d'exhaustion, qui peut virer parfois à la saturation : *Tentative d'épuisement d'un lieu parisien* (1975), *Tentative d'inventaire des aliments liquides et solides que j'ai ingurgités au cours de l'année mil neuf cent soixante-quatorze* (1976), *J.R. : tentative de saturation onomastique* (1981). Enfin, la profession de Perec, documentaliste pendant vingt ans dans un laboratoire du CNRS spécialisé en neurophysiologie médicale, lui avait donné une longue pratique du dépouillement systématique de revues et d'ouvrages, du classement et du listage, ainsi qu'une familiarité avec la terminologie scientifique.

Cette capitalisation lexicale a donc non seulement été réinvestie dans ses écrits, mais de surcroît mise en scène de plus en plus résolument au fil de l'œuvre : Perec adopte ce qu'il ne serait pas excessif de définir comme une posture de lexicographe, en ce sens que la narration accueille un métadiscours qui vise à préciser le sens, déterminer une acception, voire fournir tout simplement une définition. Dans *La Vie mode d'emploi*, on peut relever 24 passages où le texte commente, d'une manière ou d'une autre, le mot qu'il utilise ; il faut y ajouter un chapitre complet du roman, le LX, dont le héros est un lexicographe d'un genre un peu particulier, travaillant pour le compte des dictionnaires Larousse. Les motivations et les modalités de ces interventions métalinguistiques peuvent être classifiées en trois catégories principales.

1.1 L'élucidation sémantique

Dans une véritable intention dictionnairique, Perec propose parfois un mot rare dont il estime qu'il ne pourra pas être compris par le lecteur. Il peut s'agir d'un terme étranger, d'un emprunt à du vocabulaire technique ou d'un terme historiquement daté :

> Ursula Von Littau fut hissée sur un *atatich*, cette sorte de tente ronde que portent les chameaux, et dans laquelle se tiennent ordinairement les femmes […]. (Perec, 1978 : 1130)

> Sur le chambranle de la porte est accrochée une *mezouza*, ce talisman d'appartenance orné de trois lettres […]. (Perec, 1978 : 1171).

Dans les deux cas, les mots sont empruntés directement à l'arabe et à l'hébreu. Ils ne sont attestés ni dans le *TLF* (1971-1994), ni dans le *Grand Larousse de langue française* (1971), ni dans le *Petit Larousse* (editions de 1968 et 1975), ni dans le *Lexis* (1975), pas plus que dans *Dictionnaire alphabétique et analogique de la langue française* de Robert (1959), le *Petit Robert* (édition de 1977) ou le Littré (1873), tous dictionnaires que l'écrivain aurait pu consulter en son temps. Perec a pu avoir de la *mezouza*, objet connu dans la tradition cultuelle juive, une connaissance personnelle, encore qu'il ne fût absolument pas pratiquant. En revanche, il a utilisé le terme *atatich* – que le *Supplément au dictionnaire arabe* de Dozy donne pour « palanquin de parade » – , au pluriel en le construisant comme un singulier (celui-ci est en réalité la forme *attouch*), signe que l'écrivain n'hésitait pas à puiser dans le réservoir de langues étrangères qu'il ne pratiquait pas. L'introduction de ces xénismes participe d'abord d'un certain effet de réel : dans les deux cas, les mots renvoient soit à la situation géographique du personnage (Ursula von Littau est prisonnière à Alger), soit à un trait définitoire de son identité. Dans le cas de Cinoc, le geste lexicographique accompagne l'intention narrative : de la même manière que le xénisme, entraînant des italiques, *exhibe* le terme dans le texte, Cinoc, en accrochant chez lui une *mezouza*, affiche ses origines. Ce qui renvoie sans ambiguïté à une problématique constitutive de la poétique perecquienne, celle de la dissimulation, et en particulier la dissimulation de l'identité juive, à laquelle il a été confronté durant la guerre.

Ailleurs, la démarche d'élucidation mise en œuvre outrepasse son rôle classique, lorsqu'elle s'applique à des néologismes, comme ceux qui désignent les projets d'invention d'un personnage appelé Carel Van Loorens :

> Un tissu baptisé « pellette », sorte de cuir artificiel, fait d'une armure de toile forte enduite d'un mélange de liège en poudre, huile de lin, colles et résines ; ou une « forge solaire » consistant en un assemblage de plaques de métal polies comme des miroirs convergeant sur un foyer topique. (Perec, 1978 : 1126)

Là encore, ces termes ne sont pas attestés dans les grands dictionnaires de langue du XIX^{e} et du XX^{e} siècle. S'il s'agit bien, comme on peut le soupçonner, d'un néologisme perecquien, le geste lexicographique vient ici achever de renverser le rapport fiction / réel. Dans certains cas, on pourrait même parler de « pseudo-définition » : ainsi Perec amène-t-il dans le texte, sans précaution ni avertissement typographique particuliers, le terme *topazine*.

> Gormas a été empoisonné avec une quantité microscopique de topazine qui se trouvait sur la pointe du fleuret de Michard. (Perec, 1978 : 857).

Ce terme n'est attesté dans aucun des principaux dictionnaires de langue cités plus haut. Et pour cause : en avançant dans le récit, on se rend compte que le nom de cette substance (qui n'est « active que sur les individus ayant eu récemment une hépatite virale ») a été créée *ad hoc* par un groupe d'adolescents apprentis-romanciers, qui multiplient les complications narratives invraisemblables. La posture du lexicographe est ici biaisée, puisqu'elle utilise une structure explicative — le geste définitoire —, par ailleurs fonctionnelle dans le roman, pour fournir une explication qui relève de la fiction sémantique : elle rend « parfaitement crédible des savoirs fictifs » (Magné, 1997 : 30). Elle invite donc le chercheur à interroger le vocabulaire perecquien, en particulier les termes rares, avec une attention particulière, et en prenant toujours en compte le fait que le sens proposé peut faire l'objet d'un piège lexicographique.

1.2 La définition par synonymie

A l'inverse, il est fréquent que Perec utilise un mot courant, qu'il assortit ensuite de son synonyme, pouvant être soit un terme rare, soit un terme issu du vocabulaire technique. Dans ces cas, la définition fonctionne à l'envers, puisque l'explication précède la vedette :

> une de ces pendules ouvragées appelées cartels (Perec, 1978 : 828)

> Puis [Winckler] enduisait toute la surface d'un vernis protecteur qu'il appliquait avec un de ces pinceaux larges et plats appelés queue-de-morue (Perec, 1978 : 903).

> une silhouette de lion qu'en termes d'héraldique on aurait qualifié de *passant* ou de *léopardé* (Perec, 1978 : 894).

Le geste n'est pas motivé par un désir d'aplanissement de la difficulté sémantique, mais ressortit plutôt au désir de « faire concurrence au dictionnaire » (Magné, 1997), ce qui passe par l'adjonction de termes explicitement – et métalexicographiquement – désignés comme relevant d'un domaine de spécialité (peinture, horlogerie, héraldique). Ce mouvement peut être lu comme la marque du désir de saturation, évoqué précédemment, qui va passer par l'exhaustion synonymique. Mais on peut aussi l'interpréter comme une intention esthétique, que l'on pourrait rapporter au geste pictural de la miniaturiste Marguerite Winckler, l'un des personnages du roman. Perec travaille la langue dans ses « presque imperceptibles détails » (Perec, 1978 : 297), faisant de chaque mention l'occasion d'une ciselure lexicale qui transforme l'énumération (marque stylistique prégnante chez lui) en variations presque infinies.

1.3 La métalexicographie

Perec se livre également à une réflexion d'ordre métalexicographique lorsqu'il décrit le mot dans son rapport à la définition qui peut en être donnée ; il interroge alors la langue en diachronie, mettant en évidence la labilité historique de la distribution du sens.

> Les raisons de cette polémique étaient loin d'être claires pour ce lecteur du XX^{e} siècle qui ne pouvait se figurer qu'approximativement ce qu'avaient représenté les théories de Galien et pour qui des termes comme « atomistes » ou « matérialistes » n'avaient certainement plus le sens qu'ils avaient eu pour son lointain ancêtre. (Perec, 1978 : 1253-1254)

La relation du mot à sa définition, ainsi que la question de la relation entre le référent, le signifiant et le signifié se transforme dans certains chapitres en enjeu narratif, notamment lorsqu'elle se problématise. Nous citerons deux exemples : celui du personnage de Marcel Appenzzell, anthropologue parti étudier la peuplade des Orangs-Kubus. Il constate chez eux une déperdition importante du vocabulaire, liée à une tradition cultuelle (supprimer un mot lorsque quelqu'un meurt) qui aboutit à des spectres de définition invraisemblablement larges : « Ainsi *Pekee*, le mot malais désignant la chasse, voulait dire indifféremment chasser, marcher, porter, la lance, l'antilope, le cochon noir » (Perec, 1978 : 144). L'autre personnage, appelé Cinoc, est un lexicographe des dictionnaires Larousse, qui exerce le « dur métier de tueurs de mots » (Le Tellier, 2006 : 142). Son travail consiste à « éliminer tous les mots et tous les sens tombés en désuétude » (Perec, 1978 : 347). Mais pour contrer cette entreprise, Cinoc rassemble pour son compte 8 000 de ces mots rares ou oubliés, et le roman propose un extrait de « son » dictionnaire, sous la forme de trente définitions. Sur le plan métatextuel, on voit comment les deux mouvements de déperdition décrits sont l'antithèse du projet de Perec dans son rapport avec la langue, qui est au contraire une pulsion totalisante, où la réactivation sémantique joue un rôle de premier plan. Et parce que l'auteur pose un geste lexicographique intégré dans sa création, il invite la critique à en faire un en retour : la voie du dictionnaire, pour appréhender un « monstre » littéraire comme le corpus perecquien, semble être un passage obligé, tant sur le plan de la sémantique que sur celui de la poétique.

2 Pourquoi un dictionnaire Perec ?

2.1 Qu'entendre par « Dictionnaire Perec » ?

Il existe actuellement un certain nombre de dictionnaires d'auteurs, dont une série en particulier est proposée par les éditions Champion. Le terme « dictionnaire » qui entre dans leur titre ne doit toutefois pas masquer que l'intention première n'est pas lexicographique. Ces ouvrages, qui sont parfois des sommes monumentales, sont en réalité des index notionnels ou thématiques, proposant parfois de véritables articles de fond. Ils peuvent définir des éléments disparates : si l'on prend l'exemple du *Dictionnaire Sartre*, l'on y trouve aussi bien des aspects biographiques (article « Enfance »), que des titres d'ouvrages (article « *Critique de la Raison dialectique* »), des notions (articles « Empathie »,

« Histoire »), des lieux où s'est rendu l'auteur (« Maroc »), des biographies d'auteurs, voire de critiques sartriens (article « Michel Rybalka »), ou des événements historiques (« Mai 68 ») (Noudelman et Philippe, 2004). Mais en fonction de la sensibilité des équipes des rédacteurs, la part lexicographique de la « définition » est minorée, voire absente. Un tel outil, dont il ne s'agit ici de contester ni la pertinence ni l'utilité, ne nous semblait pas adapté à notre projet, qui souhaitait examiner plus spécifiquement la manière dont Perec avait affronté la langue pour mener à bien des entreprises littéraires souvent extrêmes. Il s'agissait de trouver un entre-deux satisfaisant entre une approche strictement littéraire (le caractère notionnel ou thématique ayant déjà été largement déchiffré par les ouvrages détaillés en introduction), et une pure entreprise lexicographique, qui aurait laissé de côté des questions de poétique indispensables à la compréhension de l'œuvre. Les travaux développés par Esther Lothammer (Lothammer, 2005) sur le lexique de J.-B. Pontalis nous ont convaincus que cette démarche était viable et pouvait être appliquée avec profit à un autre auteur.

2.2 Frantext

La présence d'une partie du corpus perecquien sur Frantext (5 textes au début du projet) rendait envisageable le traitement exhaustif de l'ensemble que nous visions. Celui-ci a été défini sur la base de l'édition de la Pochothèque, à savoir les sept romans de Perec publiés de son vivant ; après discussion avec plusieurs chercheurs perecquiens, *W ou le souvenir d'enfance*, texte qui n'est pas à proprement parler un roman, mais une autofiction, y a été ajouté, car c'est dans ce livre qu'est narrée toute l'histoire familiale de l'écrivain. En revanche, il a fallu se résoudre à écarter le roman intitulé *Les Revenentes* : en effet, Perec a opéré sur le matériel lexical une récriture phonétique si spectaculaire (contrairement au cas de *La Disparition*), avec effet de défiguration de la « lengge » qu'elle rend impossible son traitement automatisé, en particulier pour les opérations de lemmatisation. Les œuvres ont été numérisées au fur et à mesure des besoins, faisant passer le corpus général Perec à 11 textes, et celui du dictionnaire à 8. Ce dernier forme un ensemble de 518 000 mots, 59 897 formes et 24 206 lemmes.

L'ampleur de ce corpus rendait de toute façon irréaliste un traitement manuel, ne serait-ce que pour des raisons de localisations multiples des lemmes. Une modification sur le formulaire de constitution de corpus de Frantext a permis une sélection partielle, celle des huit romans concernés, en plus de l'affichage des lemmes en contexte, indispensable pour mener à bien la désambiguïsation des formes, ainsi que de la possibilité d'extraction de listes, dont nous détaillerons le traitement plus bas. D'autres œuvres ont également été numérisées ou sont en voie de l'être, notamment des essais et des entretiens : leur base lexicale n'est pas prise en compte dans la nomenclature du dictionnaire, mais tout rédacteur se réserve le droit d'y faire appel pour préciser une acception.

2.3 Idiosyncrasie

Perec est un auteur original, y compris au sein du groupe auquel il appartenait, l'Oulipo ; il a pratiqué tous les genres (roman, théâtre radiophonique, essai, poème) et parcouru la gamme qui va du burlesque (*Quel petit vélo à guidon chromé au fond de la cour ?*) au tragique (*W ou le souvenir d'enfance).* Chacune de ses œuvres a, en quelque sorte, réinventé sa langue, et cette hétérogénéité, pour ne pas dire cette idiosyncrasie, justifie la mise en place d'un outil critique panoramique, qui peut s'appuyer sur une base méthodologiquement susceptible d'échapper, au moins en partie, à une sectorisation critique. Comme nous l'avons signalé, une partie importante du travail perecquien recourt à un vocabulaire qui n'est pas d'usage fréquent : termes techniques relevant de la botanique, biologie, chimie, navigation, physique, terminologie historique ou liée à un contexte socio-historique (*bal-à-jo*, *harki*, *Rexistes*), archaïsmes (*vélocimane)*, xénismes (*pacemaker, skrimshandher*). D'autre part, la présence d'un volumineux cahier des charges pour *La Vie mode d'emploi*, et ses vingt et une paires de listes d'éléments imposés, a conduit l'auteur à des déclinaisons lexicales importantes sur des sèmes donnés : par exemple, l'item *volumes* se distribue en *cube*, *parallélépipède rectangle*, *cylindre*, *sphère*, etc. Parfois, le roman outrepasse la contrainte en proposant une véritable liste, qui est aussi une déclinaison sémantique. Ainsi le polyèdre imposé par *volumes* au chapitre 96 se décline-t-il en :

> modèles de cristallographie, pièces de bois minutieusement taillées reproduisant quelques formes holoèdres et hémièdres des systèmes cristallins : le prisme droit à base hexagonale, le prisme oblique à base rhombe, le cube épointé, le cubo-octaèdre, le cubo-dodécaèdre, le dodécaèdre rhomboïdal, le prisme hexagonal pyramidé. (Perec, 1978 : 1251).

Par ailleurs, Perec enchâsse, et surtout encrypte dans le texte un contenu biographique à l'aide d'allusions, de jeux phonétiques et de doubles sens, ce qui va influer sur le sémantisme de certains lemmes, en apparence anodins. L'écrivain fait en effet partie de ceux dont la destinée individuelle a croisé la grande Histoire pour le pire : enfant de Juifs polonais installés en France, Perec a d'abord perdu son père pendant la débâcle, puis sa mère Cyrla, contrainte de porter l'étoile jaune, puis raflée, et disparue quelque part entre Drancy et Auschwitz en 1943. L'enfant, devenu orphelin, a vécu caché pendant la guerre et a été recueilli par sa famille paternelle après-guerre. Le lien entre cette biographie douloureuse et une œuvre littéraire parfois perçue à tort comme essentiellement ludique (Kauffmann, 1997 : 51) est tout sauf linéaire. Certains lemmes vont fonctionner comme des *motifs*, d'autres devenir intrinsèquement porteurs de sens liés à l'histoire personnelle, au moyen d'*aencrages* (Magné, 1999 : 29). Dans un tel contexte, les termes *disparition* (titre du célèbre roman ne comportant pas la lettre *e*), *manque* (nom de l'une des contraintes du cahier des charges de *La Vie mode d'emploi*), tout le lexique de la famille, représenté par plus de trente termes (*père, mère, tante, oncle, cousin, arrière-grand-père…*), celui qui se rapporte à la dispersion et à la recomposition (*puzzle*, *fragment*, *tisser*, *coudre*, *colle*), le vocabulaire historique (*rafle)*, plusieurs nombres liés à des dates clefs (*11*, *43*) méritent donc d'être interrogés avec un soin particulier.

3 Problématiques de la définition

3.1 Constitution des listes

Frantext a permis l'établissement d'une première série de listes, qui trient l'ensemble du vocabulaire du corpus par ordre alphabétique. Ces listes présentaient toutefois un double inconvénient : elles classaient des formes non lemmatisées, et elles ne conservaient pas systématiquement la majuscule. En effet, les textes de Perec, dont la numérisation s'est échelonnée sur plus de vingt ans, n'ont pas tous été balisés suivant la même norme, ce qui crée des disparités au plan du traitement graphique lors de la récupération des informations. Jean-Luc Benoit a pallié cette difficulté en utilisant le logiciel gratuit *Dico*, developpé par Jean Véronis, qui conserve les majuscules, pour effectuer une seconde extraction du lexique ; celle-ci a ensuite été lemmatisée grâce à Cordial Analyseur 13. J.-L. Benoit a ensuite opéré une première désambiguïsation manuelle des formes homographes (Benoit, 2007), qui portait essentiellement sur des aspects syntaxiques (participes passés du type *allée*, formes conjuguées comme *ai*), formes comptabilisées comme des substantifs. Malgré le soin apporté à l'opération, il faut rester conscient des limites sémantiques parfois intrinsèquement floues des polysémies (*histoire*, *mode*, *homme*). De plus, un lecteur détectera certains homonymes homographes là où un autre ne les verra pas. L'opération de séparation des acceptions étant coûteuse en temps, nous avons pris le parti, dans la plupart des cas, de renvoyer le rédacteur, dans l'hypothèse où il désire traiter tel ou tel lemme, directement à Frantext. En permettant un affichage du terme en contexte, l'interface permet à son utilisateur d'effectuer une désambiguïsation fiable et rapide, qui peut être utilisée pour affiner le premier traitement : ainsi l'article *Divan/sofa/canapé*, consacré au mobilier, exige-t-il, pour le dernier des trois lemmes, une mise à l'écart des quatre acceptions de *canapé* comme « tranche de pain de mie taillée en rectangle, frite ou grillée, dont l'épaisseur et la grandeur varient suivant le mets qu'elle doit supporter » (*TLFi).*

Néanmoins, les problèmes divers soulevés par cette opération de désambiguïsation nous ont convaincus d'entreprendre une deuxième campagne de relecture et de nettoyage des listes. C'est ainsi qu'il a fallu statuer sur le sort des graphies en ancien ou moyen français, de type *animaulx* (qui a été rattaché à *animal*), identifier des séquences de lettres correspondant à des abréviations (*av*) et indiquer à quelle lemme elles se rattachaient, repérer les mots étrangers (*decree*) insérés dans des passages eux-mêmes en

langue étrangère. Il a également fallu identifier les lemmes inconnus appartenant à des séquences intertextuelles dont l'apparition est réglée le cahier des charges de *La Vie mode d'emploi*. L'écrivain a par exemple utilisé une série de passages empruntés à Rabelais, mais attribués dans le roman à des écrivains imaginaires (chapitres 4, 54, 64, entre autres). Certains comportent des néologismes, comme le *tarande*, animal fabuleux décrit dans le *Quart Livre* (Perec : 1975, 671) ; d'autres reprennent une séquence de *barragoin* tirée de *Pantagruel*, formée par une suite de barbarismes inidentifiables : « Al baridim gotfano dech min brin alabo dordin falbroth ringuam albaras » (Perec : 1975, 1048). Il s'agissait pour nous de vérifier que ces formes correspondaient bien à une réalité textuelle, et non, pour les plus brèves d'entre elles, à une césure intempestive lors du traitement des listes. Plus complexe est le cas de certains mots à trait d'union, voire de certaines collocations, qui, sur le plan sémantique, demandent parfois à être recomposés pour être analysés : *bow-window*, *miroir de sorcière*. Ce dernier cas présente quatre occurrences, dans trois romans différents : (Perec, 1965 : 61 ; Perec, 1978 : 691, 1357 ; Perec, 1979 : 1415). L'expression désigne un petit miroir cylindrique aux contours richement ornés, donnant l'illusion qu'il s'agit d'un œil enchâssé dans une armature. Chez Perec, sa mention renvoie, notamment dans *Un cabinet d'amateur*, à l'esthétique du leurre, du reflet, mais l'objet, ambigu, peut aussi susciter convoitise, dédain ou répulsion. On peut donc concevoir que les rédacteurs repèrent d'entrée de jeu la collocation et la signalent par une annotation dans la liste, de sorte à proposer, dans l'entrée *miroir*, une sous-entrée spécifique *miroir de sorcière*.

Enfin, et bien qu'ils n'entrent pas, sauf cas d'espèce, dans le champ couvert par le *Dictionnaire Perec*, nous avons mis à profit cette seconde campagne de relecture pour harmoniser les graphies d'une série de noms propres, qui présentaient parfois des discordances. Ces dernières pouvaient être liées à des erreurs commises par Perec lui-même, notamment au moment de la reprise de certains noms propres dans l'index, ou à des différences philologiques entre les premières éditions de *La Vie mode d'emploi* et celle, plus récente, de la Pochothèque, qui a intégré une série de corrections. C'est ainsi que les deux patronymes *Appenzell* et *Appenzzell* (deux occurrences chacun) ont été rattachés à la forme *Appenzzell* (31 occurrences), après que le rédacteur eut d'une part vérifié qu'il s'agissait bien du même personnage, et qu'il se fut d'autre part référé à l'orthographe du fac-similé du manuscrit repris dans l'édition du cahier des charges du roman. Ces modifications ont systématiquement été annotées par les rédacteurs qui les ont opérées. Même si elles concernent moins directement notre entreprise lexicographique, elles semblent essentielles à l'obtention d'un chiffrage exact du nombre de lemmes du corpus. De plus, dans l'hypothèse d'une mise à disposition de ces listes à l'usage d'autres chercheurs, il nous importait de tendre d'emblée à une réduction maximale du bruit lié aux problèmes de lemmatisation et aux erreurs typographiques.

Les listes ont ensuite été formatées sous Excel. L'ensemble comprend vingt-sept fichiers (un par lettre + un fichier qui regroupe ponctuation et signes diacritiques et typographiques). Pour chaque lemme est donnée la fréquence, distribuée (grâce à un tri croisé) ouvrage par ouvrage, avec un total récapitulatif. Les listes sont présentées par ordre alphabétique dans l'onglet principal, et par ordre décroissant des fréquences dans le deuxième onglet, de sorte que le rédacteur puisse facilement repérer les mots les plus utilisés, ainsi que ceux qui apparaissent en hapax. Chaque rédacteur se voit affecter une colonne du tableau : s'il estime que le lemme mérite définition, par lui-même ou un autre rédacteur, il entre le chiffre 1. Une colonne programmée avec une somme automatique additionne ensuite le score des « cellules-rédacteur », chiffre qui s'échelonnera entre 0 (lemme non choisi) et 4 (lemme choisi par l'ensemble des rédacteurs). La colonne de la somme automatique, retriée avec une clef de tri décroissante, permettra ainsi de savoir quels lemmes sont les meilleurs candidats à la définition.

Le choix par lecture exhaustive des listes étant une tâche ardue, nous avons imaginé de le croiser avec une autre approche. Plusieurs des rédacteurs ont constitué, au fil de la lecture, des listes de termes leur paraissant devoir figurer dans le dictionnaire. Des groupements thématiques (outils d'écriture, iconographie, mots rares, jeux…) ont guidé l'établissement de ces listes, qui peuvent être conséquentes (800 lemmes pour *La Vie mode d'emploi*, par exemple). Ainsi, en reportant ces résultats, en examinant les zones de haute et basse fréquence, et enfin en relisant la totalité des listes, on peut imaginer que le dictionnaire aura su, à terme, opérer un choix pertinent parmi les quelque 24 200 lemmes utilisés par Perec.

3.2 Nomenclature et hyperonymie

L'établissement de la nomenclature est une question complexe : comme le soulignent Collinot et Mazière, (1997 : 54), la nomenclature d'un dictionnaire de langue est « toujours en décalage par rapport au potentiel de la langue », et d'une certaine manière, cette observation vaut pour un dictionnaire d'auteur. Le premier geste a donc été de constituer une équipe de quatre rédacteurs et d'un consultant en génétique, tous chercheurs en littérature. Trois des rédacteurs (B. Magné, M. Rémy, V. Montémont), ainsi que le consultant (Ph. Lejeune), sont issus de traditions critiques différentes, mais connaissent l'œuvre perecquienne, sur laquelle ils ont préalablement travaillé, parfois à un degré de spécialisation très poussé ; la quatrième rédactrice (I. Turcan) est une lexicographe non spécialiste de Perec, qui a guidé la structuration des définitions, et qui traitera plus spécifiquement des mots rares. L'équipe compte enfin un ingénieur (J.-L. Benoit), responsable de l'établissement informatique des listes. Le dictionnaire restera donc sans doute marqué par les subjectivités de ceux qui l'auront écrit, mais on peut espérer que leurs approches différenciées sauront éviter le piège de la clôture dans un certain type de présupposé critique.

Le premier critère de choix a été quantitatif : compte tenu des contraintes liées à une édition papier, il a été convenu de ne retenir qu'un lemme sur dix, soit environ 2 500. Les critères retenus ont été détaillés en 2.3 : soit ils sont d'ordre sémantique (mot rare, néologisme, xénisme), soit ils s'appuient sur le fait que le mot joue un rôle spécifique dans la poétique de Perec, envisagée sous tel ou tel de ses aspects.

Les noms propres ont été laissés de côté : d'une part en raison de leur nombre, trop important, d'autre part parce qu'ils auraient conduit assez naturellement à l'établissement d'un dictionnaire des personnages, ce qui ne va pas dans le sens de notre démarche, ancrée dans le lexicographique, et non dans le narratif. En revanche, quelques noms propres pourront trouver leur place dans le dictionnaire, si l'onomastique est manipulée avec une intention sémantique : ainsi apparaît dans *La Vie mode d'emploi* le personnage d'un certain *George Bretzlee*, romancier américain. Son nom est le quasi-anagramme de *bretzel* ; or le patronyme *Peretz*, dont a dérivé celui de *Perec* – car les orthographes des noms juifs variaient souvent au moment de l'obtention de nouveaux papiers –, désigne selon l'auteur « en hongrois (à Budapest, plus précisément) […] ce que nous appelons "Bretzel" » (Perec, 1975 : 51). L'écrivain a donc réintroduit, avec un jeu homophonique, l'une des contraintes du cahier des charges de *La Vie mode d'emploi*, qui consiste à faire une autocitation à chaque chapitre (en l'occurrence une allusion à son premier roman, *Les Errants*) : mais au-delà, *Bretzlee* renvoie à un autobiographème, celui de l'identité, les vacillations onomastiques suggèrant que l'errance peut exister dans la langue comme elle a existé dans l'histoire individuelle. C'est pourquoi *bretzel* fera l'objet d'une entrée dans le dictionnaire, où figurera aussi le nom propre.

Il semble également important de faire jouer la notion d'hyperonymie et d'hyponymie dans les définitions. Le texte de Perec présente un coefficient de variabilité du vocabulaire « anormalement » élevé : comme le souligne Le Tellier (2006 : 246), « *La Disparition* est le roman de la langue française contenant le plus grand nombre de substantifs différents ». Le fait résulte d'une part des contraintes lipogrammatiques, qui ont conduit l'auteur à puiser dans des réservoirs lexicaux marginaux, d'autre part à l'exploration méthodique de réseaux sémantiques, notamment sous l'impulsion du cahier des charges.

Dans le dictionnaire, le regroupement sous une vedette hyperonymique permettra d'éviter l'entropie, en faisant apparaître le maillage isotopique qui surdétermine l'emploi de certains termes. Ainsi, *divan*, *sofa* et *canapé* ont-ils été regroupés sous la même définition. En effet, « Divan, canapé » est le neuvième item de la quatorzième liste du cahier des charges de *La Vie mode d'emploi* : comme *sofa*, les deux termes renvoient, sur le plan dénotatif, à la sphère domestique. Ces pièces de mobilier, qui ont des usages presque similaires, fonctionnent d'abord comme des marqueurs sociologiques : leur forme, leur revêtement, leur état d'usure renseignent sur la catégorie socioprofessionnelle, le type d'habitat. Ils définissent aussi l'horizon idéalisé de la consommation : « ils eurent envie de moquettes, de tables, de fauteuils, de divans » ; « Ils rêvaient, à mi-voix, de divans Chesterfield » (Perec, 1965 : 72). La distribution des occurrences (21 pour *canapé*, 7 pour *sofa* et 45 pour *divan*) dans le corpus montre cependant une nette prévalence de *divan*. Celle-ci s'explique en partie par le fait que *divan*, qui ne comporte pas de *e*, est employé préférentiellement dans *La Disparition*. Mais son score élevé appelle une

mise en relation avec le passé psychanalytique de Perec, qui a suivi trois analyses (dont l'une avec Dolto et l'autre avec Pontalis). Ici, c'est la comparaison des fréquences qui pourra éclairer cette dimension, liée à la biographie, qui n'est jamais explicitement évoquée dans le texte romanesque.

Autre exemple, celui du mot *plastique*, qui revient 31 fois chez Perec : il se décline, en outre, sous diverses formes chimiques, dont certaines développées après-guerre : *bakélite*, *celluloïd*, *ébonite*, *formica*, *galalite*, *polystyrène*, *skaï*, *vinyle*, *nylon*. Cette diversification, en soi intéressante, peut donner lieu à une lecture sociocritique : le plastique est parfois chez Perec le matériau du simili, du faux, du toc (celui des objets des classes moyennes qui tentent de singer les riches) ; dans d'autres cas, il est au contraire la manifestation d'une modernité agressive, et exhibée par la classe possédante, comme le montrent les « panneaux de contreplaqué habillés de skaï blanc » (Perec, 1978 : 1086) de la salle à manger dernier cri de Madame Moreau. En tout cas, sa valeur esthétique est à mettre en rapport avec un facteur économique, que la définition globale, parce qu'elle pointe une isotopie, est mieux à même de révéler qu'une collection de définitions isolées.

3.3 Structure de l'article

Une fois les lemmes choisis, il a fallu créer une structuration de l'article, susceptible de s'adapter à une triple exigence : une fonction lexicographique ; une prise en compte des spécificités de l'écriture de l'auteur ; un alignement sur les balises descriptives de la TEI (opéré par Jean-Luc Benoit), que nous nous proposons d'adapter pour cette occasion. L'article doit remplir un rôle d'élucidation sémantique, notamment dans le cas des nombreux mots rares, et donc assumer une fonction dictionnairique, sans que celle-ci soit toutefois son unique vocation. En effet, il doit aussi être capable d'aller au-delà et de proposer des éléments de stylistique, de psychanalyse, d'architecture générale de l'œuvre, de poétique. Il se structure en trois parties essentielles :

• L'identification et la description du lemme, qui comprend la vedette, la nature grammaticale, la distribution des fréquences dans le texte, œuvre par œuvre, et la distribution par nombre grammatical (occurrences au singulier *vs* occurrences au pluriel). Puis une brève définition lexicographique générale, au sens de « paraphrase sémantiquement équivalente » (Chaurand, 1990 : 276) est élaborée à partir du *TLFi* et du *Grand Robert v2* et peut éventuellement être augmentée par le recours à un autre dictionnaire, soit contemporain, soit ancien pour les mots rares. En ce qui concerne la partition traditionnelle des domaines de la connaissance et des domaines techniques (histoire, biologie, médecine, etc.), nous nous référons à la nomenclature du *TLF*, mais abrégée, de sorte à rester à un niveau de granularité élevé. L'utilisation de ces marqueurs lexicographiques « classiques » n'a pas à être systématique : on y recourra surtout pour traiter les termes polysémiques ou des mots rares.

• La définition « spécifique », qui affine l'analyse en examinant la charge sémantique particulière que Perec a pu conférer à certains termes, ou la manière dont ceux-ci participent à l'économie générale de l'œuvre. Pour clarifier l'organisation de cette rubrique, nous nous sommes appuyés sur quatre grands axes dégagés par Bernard Magné à partir des observations de Perec : le sociologique, l'autobiographique, le formel et le romanesque (Magné 1999 : 21). Ceux-ci ont été transformés en balises aux sens parfois plus étendus : SOC regroupe ainsi sociologie, sociologique, et sociocritique ; BIOG, la biographie et l'autobiographie. Nous en avons ajouté d'autres, comme CDC pour *cahier des charges*, ou encore AENC pour identifier clairement les aencrages. Ces marqueurs demeurent toutefois en nombre réduit et certains sont volontairement polysémiques et englobants, pour permettre leur application à des notions très vastes, comme la contrainte. Une liste récapitulative des différents marqueurs est proposée dans le tableau 1.

• La troisième partie de l'article, linguistique, recense les informations liées aux variantes orthographiques, à l'étymologie ou à la morphosyntaxe. Par exemple, un mot comme *sofa*, avec ses deux orthographes concurrentes, va être commenté sur ce point : « Perec utilise les deux orthographes *sofa* et *sopha* ; la seconde est exclusivement cantonnée à *La Disparition*. L'archaïsme graphique peut être une manière d'insister sur la connotation orientalisante, ou de compenser par la variété graphique l'appauvrissement lexical induit par la contrainte. » (article-test de *divan*, *sofa*, *canapé*).

Tableau 1. Liste récapitulative des marqueurs lexicographiques propres au *Dictionnaire Perec*.

AENC	= aencrage
BIOG	= biographie et autobiographie
CDC	= cahier des charges (s'applique à tous les termes inscrits dans le CDC)
COOC	= co-occurrences, dans le cas où leur étude se révèle pertinente
COLLOC	= collocation. Les collocations, en linguistique, sont des rapprochements de termes non fortuits, qu'on peut apparenter aux expressions figées
DIST	= distribution. Intervient en début de définition, pour signaler la répartition des variations morphologiques (pour les noms et adjectifs)
ETYM	= étymologie
[FORM]	= formel, contrainte
HOM/SYN	= homonymie / synonymie
ITX	= inter et intratextualité
[MOT.]	= motif (au sens morphosyntaxique : un syntagme répété)
MORPH/SYNT	= morphosyntaxe. A utiliser pour signaler les modifications morphologiques, les usages particulier ou récurrent… La néologie, par exemple, peut être rangée au nombre de ces phénomènes.
ORTH	= orthographe
ROM	= roman, romanesque
SOC	= société, sociologie, sociocritique

Les marqueurs réservés au *Dictionnaire Perec* n'ont pas à être utilisés simultanément, ni multipliés, et on peut même envisager qu'un article court ne fasse pas appel à eux. Ils ont surtout pour objectif d'orienter rapidement et efficacement la lecture de l'article vers le ou les aspects de la poétique sur laquelle ledit article souhaite mettre l'accent. D'où l'absence d'ordre prédéterminé des marqueurs (sauf pour certaines, comme ORTH /ETYM, ou DIST, organiquement rattachés à une rubrique), qui laisse toute latitude au rédacteur pour hiérarchiser l'information qu'il veut délivrer.

Trois formats d'articles ont été retenus : les « gros » articles (3 500 signes maximum) concernent des mots qui présentent un intérêt spécifique au regard de la poétique ; les articles « moyens » (c. 1 600 signes) concernent des mots récurrents au spectre sémantique plus limité, ou des éléments qui n'ont pas de rôle structurel dans la narration (« cendrier », « chat ») ; quant aux « petits articles » (c. 700 signes), ils n'auront qu'une visée lexicographique, élucider les mots rares. Cet indicateur quantitatif peut souffrir des variations, mais la limitation tend à conserver l'optique lexicographique que s'est fixé le dictionnaire, en lui interdisant de se transformer en dictionnaire thématique déguisé. Ainsi le rédacteur confronté à un article trop long pourra-t-il se demander si celui-ci n'appelle pas un éclatement sur d'autres vedettes. Par exemple, on peut envisager une entrée ***jeu***, mais une autre pour ***échecs*** et une autre pour ***go***.

4 Un exemple d'article : problématique

• Avant que ne soit opérée la sélection des lemmes, la rédaction d'articles-tests est apparue comme indispensable à la validation de nos cadres méthodologiques. Nous avons choisi le mot *père*, car il fait partie du noyau autobiographique de l'œuvre. Premier problème : fallait-il traiter à part le synonyme *papa* ? En effet, il renvoie au même référent mais n'appartient pas au même registre de langue, ni ne véhicule le même degré d'intimité. Compte tenu de la disparité dans la distribution des deux formes, il nous a paru pertinent de les maintenir dans la même entrée, de manière à souligner justement la minimisation de la seconde dans certains textes romanesques : la coexistence est traitée dans la troisième rubrique, sous l'opérateur de description SYN.

• Le deuxième problème était quant à lui lié aux fréquences du terme telles que nous les ramenaient les extractions opérées par Frantext et par *Dico* : la requête « père » aboutit à 169 résultats, mais compte tenu de l'absence de prise en compte du trait d'union, ce nombre est faussé par un bruit important, lié à tous les composés (« beau-père », « grand-père », « arrière-grand-père »). Le nombre d'occurrences ambiguës justifiait l'écriture d'une grammaire Frantext, qui a permis d'éliminer 40 formes composées, grâce à une

règle d'exclusion du trait d'union. Dans ce type de cas, il est particulièrement appréciable de pouvoir disposer non seulement d'un corpus numérisé, mais en plus des fonctionnalités offertes par le moteur de recherches Stella, qui donne la possibilité de séparer en une seule requête les formes simples des formes composées.

• Le dernier problème, plus complexe, a été la prise en compte du contexte d'emploi, et en particulier de l'utilisation du déterminant. Elle a appelé à son tour un examen des voisinages, qui permet d'établir le fait qu'il existe des proximités sémantiques (père / mort), ce qui renvoie à l'une des problématiques essentielles de notre dictionnaire : la définition se fait au sein du premier cercle de la langue, puis au sein d'un deuxième, celui de la langue de l'écrivain, Tous deux ne se superposent pas exactement, comme le montre l'exemple qui suit.

Exemple d'article-test : *père*.

Père, n.m, (129 occ., s =128, pl. = 1) | **Papa**, n.m., (54 occ., s = 52, pl. 2)
DIST. **Père** : *HQD* : 8, *W* : 67, *VME* : 50, *CA* : 4, | **Papa** : *Vélo* : 3, *D* : 47, *W* : 1, *VME* : 2.
Père et *papa* obéissent à une distribution complexe. *W* condense, logiquement, 52% des occurrences du corpus, mais une partie du livre relève de l'écriture de fiction. *Papa* est omniprésent dans *D* (88,5% des occurrences), mais la contrainte alphabétique a pesé sur le choix du terme.

DEF. Homme qui a engendré, qui a un ou plusieurs enfants. *Papa* est un terme familier et affectueux qu'emploient les enfants pour parler à ou de leur père. || ***BIOG.*** Le terme touche un point nodal de la biographie perecquienne, le deuil des parents, puisque le père de l'écrivain est mort le 16 juin 1940, à la fin de la débâcle. *W* dresse un portrait du père, s'attachant à son identité (« tout le monde appelait mon père Isie »), sa description physique, son caractère (« mon père était aussi un brave à trois poils »), les circonstances de sa mort (« mon père perdit tout son sang et mourut pour la France avant d'avoir été opéré »). Perec mentionne le lieu où son père est enterré, avec le récit de la visite sur la tombe, qui est l'occasion d'inscrire dans le texte le *nom du père* : « PEREC ICEK JUDKO ». Il ressort de ces reconstitutions de souvenirs, souvent empruntés à la famille paternelle, l'image d'un être attachant, débrouillard et gai (« J'aime beaucoup dans mon père son insouciance »). || ***SYNT.*** Dans *W*, Perec privilégie la forme personnelle « mon père » (47 occurrences sur 67). Le changement de déterminant, lorsqu'il advient, correspond à un changement de point de vue : emblématisation de l'image photographique, qui suggère une transfocalisation avec l'emploi de l'article défini : « Sur cette photo, le père a l'image du père ». La deuxième personne restitue les voix familiales : (« ton père est mort ») ce qui crée un redoublement en forme d'écho. L'indéfini conjugué au passé composé marque le point où l'histoire cesse tragiquement de s'inscrire dans la banalité : « comme tout le monde, ou presque, j'ai eu un père et une mère ». *D* met en scène un père vengeur et un dispositif qui oblige tout parent ayant plus d'un enfant à le confier à « un papa adoptif » ; *VME* imagine un père vengeur, Sven Ericsson, rendu fou de douleur par la mort de son enfant, qui tue à son tour un père et une mère. || ***ROM***. Dans *W*, 5 des emplois de *père* se font dans la même phrase que le substantif *mort* ou le verbe *mourir*. L'œuvre romanesque relaie ce voisinage syntaxique : dans *VME*, le père d'Elisabeth de Beaumont se suicide, laissant une fille qu'il n'a jamais vue (C6), celui de Vera Orlova est fusillé sous les yeux de sa fille (C31). Le père du duc de Dinteville meurt dans un accident de lustre (C14), celui de Vittorio Grifalconi décède des suites d'une trichinose, celui d'Olivier Gratiolet de brucellose (C68), celui de Winckler sur un champ de bataille à Verdun (C53), ce qui n'est pas sans faire écho à la mort d'Izie Perec dans un cadre militaire. Ces différentes situations associent à la relation père-enfant l'idée de la menace, de la violence, de la défection de la structure familiale sous l'emprise de circonstances extérieures.

SYN. Excepté son emploi comme synonyme de *père* dans *D*, Perec n'utilise qu'une fois le mot « papa » : « j'aurai beau […] rêvasser pendant deux heures sur la longueur de la capote de mon papa, je ne retrouverai jamais […] que […] le scandale

de leur silence et de mon silence » (*W*, 59). Le terme ici peut être compris comme empreint d'une certaine ironie, et comporte en même temps une connotation affective, par ailleurs rare dans *W*.

5 Conclusion

Le dictionnaire Perec, s'il a eu quelques devanciers en littérature française, comme le *Glossaire des œuvres complètes* d'Apollinaire, de Claude Debon (1988), le *Dictionnaire San Antonio* (Le Dorand, Pelloud et Rosé, 1993) ou, dans un registre parallèle, le *Dictionnaire des personnages de Queneau* (David, 1994**)**, est une démarche novatrice, car elle est la première à s'appuyer sur un traitement informatisé des corpus. Les critères de choix prennent donc en compte les fréquences et la distribution, ainsi que des faits morphosyntaxiques, comme les variantes orthographiques, les voisinages, les collocations. L'intention de départ n'est pas thématique, mais bel et bien lexicographique : il s'agit de proposer un dictionnaire de la langue de Perec, qui remplisse un rôle définitoire (tout lecteur pourra y trouver la définition d'un mot rare utilisé par l'auteur), ainsi qu'un rôle critique. L'articulation des deux a été clairement mise en place grâce à une structure d'article, puis validée par une dizaine d'articles-tests dans les trois formats proposés. Les problèmes auxquels nous nous heurtons actuellement sont liés au grand nombre des lemmes à traiter, qui risquent de nous conduire à recruter deux nouvelles rédactrices.

Il est également prévu que le dictionnaire, outre sa publication papier, connaisse une vie électronique. A cette fin, Jean-Luc Benoit a rédigé une feuille de style de saisie des articles, incluant les balises spécifiques au dictionnaire, qui lui permettra une extraction rapide des données et assurera la conformité avec la TEI, en vue d'une publication de l'ouvrage sous forme de CD-Rom. En parallèle, une version web pourrait proposer les lemmes de la lettre *A*, ainsi qu'un certain nombre de matériaux thématiques qui n'auraient pas été retenus dans la définition. Nous pensons en particulier aux mots absents du corpus perecquien, alors que l'on pourrait s'attendre à les y trouver, et dont l'effacement n'est pas le fruit du hasard. Enfin, le dictionnaire de la langue perecquienne pourrait être un bon embrayeur à un dictionnaire des personnages. La question qui se posera sera de savoir dans quelle mesure la méthode développée à l'ATILF pour l'établissement du dictionnaire Perec pourra être systématisée. Si tel est le cas, il s'agira de trouver des méthodes efficaces de capitalisation et de transmission de ce savoir, de sorte à pouvoir l'appliquer à d'autres auteurs manipulateurs de langage, comme Jacques Roubaud, Raymond Queneau ou BorisVian.

Bibliographie

Benoit, J.-L. (2007). Matériaux documentaires pour le "Dictionnaire Perec" : Constitution de listes de vocabulaire, http://www.atilf.fr/perso/benoit/.

Bénabou, M. (1984). Perec et la judéité. In Magné, B. (éd.) (1984). *Cahiers Georges Perec*, *1*. Paris : POL, 15-29.

Burgelin, C. (1990). *Georges Perec*. Paris : Seuil.

Burgelin, C. (1996). *Les parties de dominos chez Monsieur Lefèvre*. Saulxures. Circé.

Chaurand, J. (1990). Quelques jalons rétrospectifs. La définition lexicographique. In Centre d'Etudes du Lexique. *La définition*. Paris. Larousse.

Collinot, A., Mazière, F. (1997). *Un prêt à parler : le dictionnaire*. Paris. PUF.

Constantin, D. (2000). "Ne rien nier. Énoncer.": la mise en place des instances énonciatives et narratives dans les premiers temps de la rédaction de *La Vie mode d'emploi* de Georges Perec. *Texte. Revue de critique et de théorie littéraire*, *27/28*, 267-296.

Constantin, D. (2004). La rédaction de *La Vie mode d'emploi* de Georges Perec : la pièce de la mémoire. *Genesis*, *23*, p. 43-55.

David, P. (1994). *Dictionnaire des personnages de Raymond Queneau*. Limoges. PULIM.

De Bary, C. (2005). L'aencrage dans la phrase perecquienne. In Beaumatin, E. et Ribière, M. (éd.) (2005). *De Perec, etc., derechef*. Nantes. Joseph.K.

De Bary, C. et al. (2006). *Cahiers Georges Perec, 9 : Le cinématographe*. Bordeaux. Le Castor Astral.

Debon, C. (1988). *Apollinaire : glossaire des œuvres complètes*. Paris. Publication de la Sorbonne Nouvelle.

Le Dorand, S., Pelloud F., Rosé P. (1993). *Dictionnaire San-Antonio*. Paris. Fleuve Noir.

Grand Larousse de la langue française. (1971). Paris. Larousse.

Le Grand Robert v2. (2005) CD-Rom, Paris. Le Robert.

Hatzfeld, A., Darmsteter, A. (1932). *Dictionnaire général de la langue française*. Paris. Delagrave.

Hartje, H, Magné, B, Neefs, J. (éd.) (1993). *Cahier des charges de La Vie mode d'emploi*. Cadeilhan-Paris, Zulma/CNRS Editions.

Hoyer, F. (1997). *Mise à l'index d'un récit : W ou le souvenir d'enfance de Georges Perec*. Mémoire de maîtrise, Université de Toulouse-Le Mirail.

Kauffmann, V. (1997). *Poétique des groupes littéraires.* Paris. PUF.

Lejeune, P. (1991). *La mémoire et l'oblique. Georges Perec autobiographe*. Paris. P.O.L.

Le Tellier, H. (2006). *Esthétique de l'Oulipo*. Bordeaux. Le Castor Astral.

Lexis, dictionnaire de langue française. (1975). Paris. Larousse.

Littré, E. (1873), *Dictionnaire de la langue française*. Paris. Hachette.

Lothammer, E. (2005). *Le vocabulaire de J.-B. Pontalis*. Mémoire de M2, Université de Nancy 2.

Magné, B. (1989). *Les Revenentes* : de l'effervescence entre lengge et texte. In Magné, B. (1989). *Perecollages 1981-1988*, 175-192.

Magné, B. (1997). Perec : faire concurrence aux dictionnaires. *Le Cabinet d'amateur*, *6*, 13-30.

Magné, B. (1999). *Georges Perec*. Paris. Nathan.

Magoudi, A. (1996). *La lettre fantôme*. Paris. Minuit.

Noudelmann, F. et Philippe, G. (2004). *Dictionnaire Sartre*. Paris. Champion.

Perec, G. (1965). *Les Choses*. Paris. Julliard, puis (2002), La Pochothèque.

Perec, G. (1969). *La Disparition*. Paris. Denoël, puis (2002), La Pochothèque.

Perec, G. (1972). *Les Revenentes*. Paris. Julliard, puis (2002), La Pochothèque.

Perec, G. (1975). *W ou le souvenir d'enfance*. Paris. Denoël, puis (1995), Tel Gallimard.

Perec, G. (1978). *La Vie mode d'emploi*. Paris. Hachette, puis (2002), La Pochothèque.

Perec, G. (1979). *Un cabinet d'amateur.* Paris. Balland, puis (2002), La Pochothèque.

Petit Larousse en couleurs. (1968). Paris. Larousse.

Petit Larousse illustré. (1975). Paris. Larousse.

Petit Robert 1. (1977). Paris. Société du Nouveau Littré.

Reggiani, C. (2001). Perec : une poétique de la photographie. *Le Cabinet d'amateur sur Internet*, http://www.cabinet-perec.org.

Rémy, M. (2003). *La concession. Pour une éthique des transformations textuelles dans l'œuvre de Georges Perec et Guy Debord.* Thèse de Doctorat. Université de Nancy 2.

Robert, P. (1959). *Dictionnaire alphabétique et analogique de la langue française*. Paris. Société du Nouveau Littré.

Robin, R. (1993). *Le deuil de l'origine*. Saint-Denis. Presses Universitaires de Vincennes.

[i] L'*aencrage*, concept élaboré en croisant la notion d'ancrage (mémoriel et géographique) et celle d'encrage (comme trace matérielle et graphique) est un procédé typique de l'écriture de Perec. Il consiste à utiliser dans le texte des éléments que B. Magné définit comme « formes-sens », et qui présentent trois caractéristiques : leur récurrence, leur lien avec un fragment d'autobiographie de l'auteur, et enfin leur utilisation comme « principe organisateur pour la forme de l'expression », en d'autres termes, comme contrainte formelle. Ainsi le *onze* est un chiffre qui est répété de manière ostensible dans *Quel petit vélo à guidon chromé au fond de la cour ?* ; il renvoie à la date, le 11 février 43, à laquelle Cyrla Perec, mère de l'auteur, a été déclarée décédée. Enfin, le onze règle le fait que le titre comporte onze mots, et que le mot *onze* soit lui-même répété onze fois dans le texte (Magné, 1999 : 29-30).